



Análisis de métodos estadísticos para evaluar el desempeño de modelos de simulación en cultivos hortícolas¹

Analysis of statistical methods to evaluate the performance of simulation models in horticultural crops

Freddy Soto-Bravo², María Isabel González-Lutz³

¹ Recepción: 2 de julio, 2018. Aceptación: 7 de enero, 2019. Este trabajo formó parte de la tesis doctoral del primer autor en Producción Vegetal en Cultivos Protegidos, Universidad de Almería, España.

² Universidad de Costa Rica. Escuela de Agronomía, Estación Experimental Agrícola Fabio Baudrit Moreno, Alajuela, Costa Rica. freddy.sotobravo@ucr.ac.cr (autor para correspondencia, <https://orcid.org/0000-0003-1959-9597>).

³ Universidad de Costa Rica, Facultad de Ciencias Económicas, Escuela de Estadística. San José, Costa Rica. mariaisabel.gonzalezlutz@ucr.ac.cr

Resumen

Introducción. Todo modelo de simulación debe ser calibrado y validado, de lo contrario las conclusiones pueden ser especulativas y erróneas. Los métodos para evaluar modelos de simulación habitualmente se aplican “por costumbre”, evitando entrar en detalles metodológicos básicos, lo cual ha sido causa de que se utilice terminología y simbología que conllevan a la confusión. **Objetivo.** El objetivo del presente estudio fue analizar los diferentes métodos estadísticos utilizados para evaluar el desempeño de modelos de simulación en agricultura, y así proponer el método más apropiado desde el punto de vista práctico. **Materiales y métodos.** Se analizaron métodos estadísticos basados en el análisis de diferencias y en el análisis de regresión, entre valores medidos y simulados. El análisis de diferencias, incluyó la raíz cuadrada del error medio estándar (RMSE), el error medio absoluto (EMA), el error relativo (ER), el índice de ajuste (d), el sesgo medio del error (MBE) y la eficiencia del modelo (E). En el análisis de regresión se analizó el intercepto, los coeficientes de regresión lineal (b) y de determinación (R^2), y los límites de confianza de la estimación. **Resultados.** El ER, el d y la E, son medidas cuyo objetivo es más la comparación entre diferentes modelos que la evaluación del desempeño de un modelo como tal. La raíz cuadrada del error medio usualmente utilizada para evaluar diferencias entre valores observados y simulados, es diferente al RMSE de una regresión. En los diferentes casos ilustrados con el modelo “Eurotate_N”, demostraron la correcta aplicación práctica del análisis de regresión como herramienta estadística para evaluar su capacidad para simular rendimiento de fruto, humedad volumétrica del suelo, evapotranspiración y materia seca del cultivo de tomate bajo invernadero. **Conclusión.** El método estadístico propuesto más apropiado para evaluar un modelo de simulación en tomate fue el análisis de regresión.

Palabras clave: evaluación estadística, modelación de los cultivos, regresión lineal, análisis de diferencias.

Abstract

Introduction. Every simulation model must be calibrated and validated, in order to avoid speculative and inaccurate conclusions. The methods to evaluate simulation models are usually applied “by habit”, without specifying



basic methodological details which leads to the use of terminology and symbology that could cause confusion

Objective. The objective in the present study was to analyze the different statistical methods employed to evaluate the performance of simulation models in agriculture, and thus propose which is the most suitable from the practical point of view. **Materials and Methods.** Statistical methods based on difference and regression analysis, between measured and simulated values were analyzed. Regarding the difference analysis group, the used methods were root mean square error (RMSE), mean absolute error (MAE), relative error (RE), adjustment index (d), mean bias error (MBE) and the model efficiency (E). In the case of the regression analysis the intercept, linear regression (b) and determination (R^2) coefficients, and the estimation confidence limits were scrutinized. **Results.** The ER, d and E, are measures which objective is the comparison between different models to simulate a given variable, instead of evaluating the performance of the model as such. The root square mean error usually used to evaluate differences between observed and simulated values is different from the RMSE regression. The different cases illustrated with the “Eurotate_N” model demonstrated the appropriate practical application of the regression analysis as statistical tool to evaluate its capacity to simulate fruit yield, volumetric soil moisture, evapotranspiration and dry matter in tomato crop under greenhouse. **Conclusion.** The most appropriate statistical method proposed to evaluate a simulation model in tomato was the regression analysis.

Keywords: statistical evaluation, crop modelling, linear regression, analysis of differences.

Introducción

El crecimiento de la población mundial (Tilman et al., 2002), asociado al cambio climático y la crisis del agua, plantean la necesidad de investigar para generar nuevas tecnologías que permitan enfrentar la creciente demanda mundial de alimentos (Galloway et al., 2004). Sin embargo, la investigación tradicional basada en experimentos de campo tiene un alto costo de inversión en infraestructura, equipo, mano de obra y tiempo. Una alternativa a la investigación convencional, son los modelos de simulación en cultivos hortícolas, que muestran una representación virtual simplificada de los procesos que ocurren en un sistema real (Yang et al., 2014), donde hay un conjunto de variables que interactúan y evolucionan mostrando un comportamiento dinámico y real en el tiempo. Estos modelos permiten la experimentación, sustituyendo la investigación tradicional basada en experimentos de campo, y permitiendo realizar una evaluación económica y rápida del efecto de diversas alternativas de manejo agrícola, reduciendo el riesgo, el tiempo de respuesta y los costos (Meinke et al., 2008). En general, las entradas (“inputs”) en la base de datos de un modelo, pueden ser diferentes a los “inputs” reales de la región donde se va a utilizar. Por tanto, para que un modelo de simulación en cultivos hortícolas, represente un proceso real, debe ser evaluado considerando las diferencias entre sistemas de cultivo, suelos, clima y prácticas de manejo, entre otros; de lo contrario las conclusiones pueden ser especulativas y erróneas (Thornley y Johnson, 2005; Cannavo et al., 2008).

El proceso de evaluación del desempeño de un modelo de este tipo, involucra su calibración y validación. En la calibración, los parámetros en la base de datos del modelo son ajustados (i) a través de una serie de escenarios de simulación hasta obtener un ajuste aceptable entre valores simulados y observados en experimentos de campo, y/o (ii) a partir de datos experimentales de la literatura para la región donde se va a validar. La validación es el procedimiento mediante el cual se evalúa el desempeño del modelo, contrastando los valores simulados de una determinada variable con datos reales obtenidos en experimentos de campo. Por ello, el objetivo principal al evaluar el desempeño de un modelo de simulación de cultivos, es valorar su utilidad práctica como herramienta de investigación y/o apoyo en la toma de decisión en aspectos de manejo y planificación a nivel de finca, regional o nacional.

Los métodos más habituales para evaluar la confiabilidad de modelos de simulación, se basan en el análisis de diferencias entre valores medidos y simulados y en el análisis de regresión, también entre valores medidos y simulados (Willmott, 1982; Yang et al., 2000; Stockle et al., 2004; Yang et al., 2014). Sin embargo, la mayoría de autores que investigan con modelos de simulación de cultivos, utilizan dichos métodos evitando entrar en detalles metodológicos básicos y utilizando terminología y simbología que conllevan a la confusión. Además, es importante identificar cuáles son los métodos adecuados cuando se compara la capacidad de simulación entre dos modelos diferentes y cuáles son apropiados para evaluar el desempeño de un modelo como tal. Por ejemplo, en el análisis de diferencias, estadísticos como el error relativo (RE), el índice de ajuste (d) y la eficiencia del modelo (E), podrían ser más útiles cuando se compara la capacidad de simulación de un modelo contra otro, pero no cuando se compara lo observado contra lo simulado en un mismo modelo. El error relativo (RE), que relaciona el error entre valores medidos y observados, respecto al promedio medido, representa el tamaño relativo de la diferencia promedio (Willmott, 1982), indicando si la magnitud de la raíz cuadrada del error medio estándar (RMSE, por sus siglas en inglés) es bajo, medio o alto. Sin embargo, tiene la desventaja que puede verse afectado por la magnitud de los valores, por valores extremos y el número de observaciones. Puede darse el caso que dos grupos de datos con valores altos y bajos, respectivamente, presenten un RMSE similar, sin embargo, al tener diferentes promedios, ambos valores de ER también serán diferentes.

El análisis de regresión por su simplicidad de aplicación, es frecuentemente usado de forma errónea para evaluar modelos de simulación. En algunos casos, el RMSE que mide la diferencia promedio entre valores medidos y simulados (Willmott, 1982), tiende a utilizarse indiscriminadamente, sin considerar que es diferente al RMSE obtenido en el análisis de regresión. El coeficiente de determinación (R^2) es una medida del grado de ajuste del modelo de regresión lineal, que cuando se usa en forma aislada no tiene sentido, ya que el objetivo de la regresión es evaluar el modelo de simulación de cultivos, no el modelo de regresión obtenido. La magnitud del R^2 no refleja necesariamente si los datos simulados representan bien a los datos observados, ya que no está relacionado consistentemente con la exactitud de la predicción (Willmott, 1982). Eso se debe a que se puede obtener un R^2 cercano a 1, pero por debajo o por encima de la línea 1:1 o identidad, teniendo una tendencia a simular valores altos o bajos que sobreestiman o subestiman, respectivamente, los valores observados.

Algunos estudios como el de Yang et al. (2014), analizaron detalladamente diferentes medidas de desviación, tales como el RMSE, el error medio absoluto (MAE), el error medio (E), el índice de ajuste (d) y la eficiencia del modelo (E).

El modelo EU-Rotate_N descrito por Rahn et al. (2010), fue creado como herramienta de apoyo para la toma de decisiones en el manejo del agua y el nitrógeno (N), en setenta especies hortícolas en rotación. El modelo simula de forma diaria, el crecimiento de la biomasa aérea y radicular del cultivo, la extracción de N, la evapotranspiración del cultivo (ETc), el volumen de drenaje, la lixiviación de nitratos, los contenidos de humedad volumétrica y N mineral del suelo, la mineralización-inmovilización del N de la materia orgánica y un análisis económico de las estrategias de manejo del N.

El modelo EU-Rotate_N es una herramienta útil para evaluar los efectos agronómicos, ambientales y económicos de las prácticas de manejo del agua y abonado de N; en la definición e implementación de leyes, planes de actuación y buenas prácticas agrarias, en el uso del N en la agricultura (Rahn et al., 2010). Además, ha sido utilizado para evaluar el efecto de las prácticas de manejo del N en rotación de cultivos al aire libre, sobre los contenidos de N mineral y humedad el suelo, la extracción de N, la producción de materia seca (MS) y el rendimiento (Rahn et al., 2010; Doltra y Muñoz, 2010). En cultivos bajo invernadero, Guo et al. (2010) y Sun et al. (2012), demostraron que EU-Rotate_N es una herramienta útil para estimar la demanda de N de pepino.

El objetivo del presente estudio fue analizar los diferentes métodos estadísticos habitualmente utilizados para evaluar el desempeño de modelos de simulación, y así proponer cuál es el más apropiado desde el punto de vista práctico.

Materiales y métodos

Descripción, comparación y crítica de los métodos estadísticos comúnmente usados para la evaluación

Los métodos usualmente utilizados en diversos estudios para evaluar el desempeño de los modelos de simulación, están basados en dos áreas: (1) en el análisis de diferencias y (2) en el análisis de regresión, entre valores medidos y simulados (Willmott, 1982; Yang et al., 2000). En el primer caso, los índices estadísticos comúnmente utilizados son (1) la raíz cuadrada del error medio estándar (RMSE) (Fox, 1981), (2) el error medio absoluto (EMA) (Willmott, 1982), (3) el error relativo (ER) (Stockle et al., 2004), (4) el índice de ajuste (d) (Willmott, 1982), (5) el sesgo medio del error (MBE) (Willmott, 1982) y (6) la eficiencia del modelo (E) (Nash Sutcliffe, 1970). En el análisis de regresión se han utilizado: (1) el intercepto, (2) el coeficiente de regresión lineal y (3) el coeficiente de determinación.

Entre los métodos de análisis de diferencias, el error relativo (ER), el índice de ajuste (d) y la eficiencia del modelo (E), son medidas cuyo objetivo es más la comparación entre diferentes modelos para simular una variable dada, que la validación del desempeño de un modelo como tal.

Análisis de diferencias entre valores observados y simulados

El análisis de diferencias se enfoca en comparar el valor calculado por el modelo de simulación (X), con el promedio de los valores observados en las condiciones definidas para el cálculo de la simulación (Y). Se propone un promedio de las diferencias (ecuación 1), al que llama "Mean Bias Error" (MBE). La varianza de las diferencias (ecuación 2) es la variabilidad alrededor de ese promedio (Fox, 1981):

$$MBE = \frac{1}{n} \sum_{i=1}^n (X_i - Y_i) \quad (1)$$

$$S_d^2 = \frac{1}{N-1} \sum_{i=1}^n (X_i - Y_i - MBE)^2 \quad (2)$$

También Fox (1981), propone el error absoluto promedio (MAE, por sus siglas en inglés) (ecuación 3), el cuadrado medio del error (MSE, por sus siglas en inglés) y su raíz cuadrada (RMSE, por sus siglas en inglés) (ecuación 4). Ellos son:

$$MAE = \frac{1}{N-1} \sum_{i=1}^n |X_i - Y_i| \quad (3)$$

$$RMSE = \sqrt{\frac{1}{N-1} \sum_{i=1}^n (X_i - Y_i)^2} \quad (4)$$

Por otra parte, Willmott (1982) sugiere que las mejores medidas para determinar el desempeño del modelo son estas dos últimas, ya que resumen las diferencias entre lo calculado por el modelo y lo que ocurre en la realidad.

Al analizar la medida estadística usualmente utilizada con el nombre de la raíz cuadrada del cuadrado medio del error (RMSE), se tiene la siguiente situación:

En el contexto del análisis de regresión (ecuación 5):

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y})^2}{n-2}} \quad (5)$$

Mientras que en el contexto del análisis de diferencias (ecuación 6):

$$\text{RMSE} = \sqrt{\frac{1}{N-1} \sum_{i=1}^n (X_i - Y_i)^2} \quad (6)$$

Evidentemente, aunque las llamen igual no son lo mismo. Ambas implican el cálculo de diferencias, pero en la ecuación 6 las diferencias se dan entre un valor ajustado por un método estadístico (Y) y el valor observado para la variable en cuestión (X); mientras que en la expresión (ecuación 5) las diferencias son entre el valor simulado y el valor real. En ambas medidas lo que se busca es obtener el valor más bajo posible, sin embargo, ambas están afectadas por la unidad de medida y el orden de magnitud de los datos. Por tanto, por sí solas no dicen mucho sobre la confiabilidad del modelo de simulación. Por consiguiente, tanto MAE como las dos medidas denominadas RMSE quedarían descartadas como estadísticos útiles para evaluar los resultados del modelo. Consecuentemente, en lo sucesivo el término RMSE se utilizará solamente en el contexto del análisis de regresión.

Si se afirma que el RMSE no es útil por sí mismo para calificar la confiabilidad del método de simulación, entonces, ¿para qué es útil? En el contexto del análisis de regresión, el RMSE mide, en promedio, cuánto se alejan los datos observados de los estimados por el modelo de regresión. Si el modelo de regresión estimado tiene intercepto 0, pendiente 1,0 (lo que se ha llamado la línea 1:1 o identidad) y el RMSE es 0, entonces se tiene el ajuste ideal y por tanto, la precisión para simular del modelo, es perfecto. En el mundo real nada es perfecto, y aunque el RMSE fuera 0, la línea de ajuste puede no ser la identidad; por otro lado, aun teniendo una línea de ajuste 1:1 el RMSE podría ser muy alto, y por ende, los datos suministrados por la simulación no serían confiables.

Lo mismo que ocurre con el RMSE en cuanto a la incapacidad de calificar por sí mismo el desempeño de un modelo de simulación, ocurre con las demás medidas propuestas. Todas las medidas que han sido planteadas para evaluar el modelo de simulación basado en las diferencias entre valores observados y simulados, calculan dichas diferencias para una muestra; con lo cual proveen información del desempeño del modelo de simulación para esa muestra específica, pero si su resultado se generaliza a la población, no se dispone de una medida del error asociado a esa generalización. Esto limita en gran medida su utilidad, por lo que en este trabajo se propone utilizar para la evaluación del desempeño del modelo, solamente el análisis de regresión, que permite la evaluación del error asociado a la generalización de los resultados, tal como se ilustra en los estudios de casos que se presentarán más adelante.

Análisis de regresión

Los modelos de simulación tienen como objetivo estimar, con base en un determinado conjunto de valores (X) que son fijos para cada condición estudiada y una relación funcional, el comportamiento de una determinada variable observada con base en una muestra aleatoria (Y). El valor que produce el modelo de simulación en este contexto es un valor dado, un valor fijo, con el cual se busca estimar el promedio de una variable a la cual se asocia una distribución y que se mide con base en una muestra aleatoria. Por tanto, el modelo de regresión a utilizar para establecer la concordancia entre el valor observado y el valor simulado es el siguiente (ecuación 7):

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon \quad (7)$$

Donde:

Y = es el valor real que se busca predecir, el valor observado bajo los parámetros (“inputs”) que se definieron en la simulación. Es una variable aleatoria con una distribución dada para cada nivel de X.

X = valor simulado por el modelo en función de los factores definidos para la simulación. Es un valor fijo, puesto que cada vez que se le den los mismos factores producirá el mismo valor. La X sólo cambiará si cambian las

condiciones de la simulación, es decir, los factores que la producen. Al respecto, Mendenhall et al. (1986), definen los valores de X como “constantes conocidas”.

ε = error aleatorio

Para cada valor obtenido en la simulación (X), existe una gama de posibles valores observados (Y) que siguen una distribución que se supone normal. También se hace el supuesto de que las variancias de Y dado X son todas iguales (homoscedasticidad).

Idealmente, para que la capacidad de predicción del modelo sea totalmente confiable, los valores obtenidos por la simulación deberían ser iguales a los valores reales obtenidos en experimentos de campo. Para esto, las condiciones ideales que deberían darse son:

β_0 , que es el intercepto de la ecuación debe ser 0.

β_1 , que es el coeficiente de regresión o pendiente de la ecuación debe ser 1,0.

R^2 , que es el coeficiente de determinación que mide el porcentaje de la variabilidad de Y explicada por cambios en X debe ser 1,0.

β_0 , β_1 y R^2 , son parámetros desconocidos.

Las condiciones anteriores reducen el modelo de regresión a un modelo que presenta la función identidad (ecuación 8):

$$Y_i = X_i + \varepsilon \quad (8)$$

Evidentemente no se dispone de todos los valores poblacionales para la variable experimental observada, por lo que esta debe medirse con base en una muestra. De igual manera, no se dispone de los valores para las β 's, ni para R^2 que deberán estimarse con base en la misma muestra. El intercepto y el coeficiente de regresión se estiman utilizando la técnica de mínimos cuadrados (Draper y Smith, 1998) y se obtiene (ecuación 9):

$$\hat{Y}_i = b_0 + b_1 X_i \quad (9)$$

El ajuste será mejor y por ende mayor la confiabilidad del modelo de simulación, cuanto más se acerquen los valores observados a los valores estimados con base en las constantes calculadas en la simulación. La diferencia ($Y_i - \hat{Y}_i$) se llama “error de estimación” para la observación “i”.

La ecuación (4) produce el promedio del error de estimación de la siguiente manera (ecuación 10):

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n-2}} \quad (10)$$

donde, RMSE es la raíz cuadrada del cuadrado medio de error de estimación (Root Mean Square error). El valor \hat{Y}_i es una estimación puntual del promedio de la variable observada. Esa estimación está sujeta a error. Para medir ese error se utilizan los límites de confianza, de la siguiente forma (ecuación 11):

$$P\left(\hat{Y}_i - t_{\alpha/2} RMSE \sqrt{\frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum_{i=1}^n x^2}} < \mu_{y/x} < \hat{Y}_i + t_{1-\alpha/2} RMSE \sqrt{\frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum_{i=1}^n x^2}}\right) = 1 - \alpha \quad (11)$$

Donde, $\mu_{y/x}$ es la media verdadera de Y dado un valor específico, $X = X_j$, en otras palabras, es la media verdadera de la variable observada en un nivel dado de la variable obtenida en la simulación, es la media de los valores observados en la simulación.

En la ecuación (5) el término \sum^2 se calcula así (ecuación 12):

$$\sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n X_i)^2}{n} \quad (12)$$

Donde n = tamaño de la muestra.

Los límites descritos pueden ser utilizados para afinar la predicción modelando el mejor y el peor escenario para una predicción dada.

En todos los casos que se presentan a continuación se cumplen los supuestos de normalidad de los errores y de homoscedasticidad.

Estudio de caso con análisis de regresión

Para ilustrar la aplicación práctica del análisis de regresión propuesto como herramienta estadística para evaluar la capacidad de simulación de un modelo, se utilizó el modelo de simulación EU-Rotate_N (EU_N), descrito por Rahn et al. (2010).

Se seleccionaron estudios de casos del trabajo realizado por Soto et al. (2014), con diferentes variables de respuesta y diferentes resultados. Los datos experimentales se obtuvieron de cuatro ciclos de cultivo de tomate (*Solanum lycopersicum* L.) en secuencia (Cuadro 1), en suelo y bajo invernadero. En dichos ciclos de cultivo, se aplicaron tratamientos de nitrógeno (N), que fueron descritos de acuerdo con: (a) la estación de cultivo (P: primavera, OI: otoño-invierno), (b) el año de cultivo: 2008, 2009, 2010 y 2011 y (c) la concentración de N en solución nutritiva (Cuadro 1).

Cuadro 1. Concentración de nitrógeno (N), N aplicado, volumen y número de riegos, N mineral en el suelo al trasplante y materia seca (MS), para cuatro cultivos de tomate (*Solanum lycopersicum* L.) y sus tratamientos, utilizados para el análisis de casos en este estudio. Almería, España. 2011-2012.

Table 1. Nitrogen concentration (N), N applied, volume and number of irrigations, mineral N in the soil when transplanting, and dry matter (DM) for four tomato (*Solanum lycopersicum* L.) crops and their treatments, used for the analysis of cases in this study. Almeria, Spain. 2011-2012.

Cultivos	Fechas de cultivo	Tratamientos de N	[N] mmol/l	N aplicado (kg/ha)	Riego (mm)	Nº de riegos	N trasplante (kg/ha)	MS (t/ha) objetivo
P-2008 (cv. Brenda)	De 10/01/08 al 21/06/08	P-08-N _c	12,0	700	418	94	542	11,7
OI-2009 (cv. Pitenza)	De 16/09/09 al 15/03/10	OI-09-N ₂	6,3	168	255	63	353	10,7
		OI-09-N ₃	12,4	320	256	63	287	10,8
OI- 2010 (cv. Razimo)	De 05/08/10 al 25/01/11	OI-10-N ₁	0,64	33	285	66	44	11,8
		OI-10-N ₂	4,4	159	270	66	64	13,3
		OI-10-N ₃	13,6	491	282	66	28	12,8
P-2011 (cv. Ramyle)	De 14/03/11 al 14/07/11	P-11-N ₁	1,1	28	174	50	17	9,2
		P-11-N ₂	5,2	151	218	50	14	11,2
		P-11-N ₃	13,4	425	240	50	13	11,7
		P-11-N ₄	21,7	677	237	50	66	11,8

En el cultivo del año 2008, el manejo del riego y la concentración de N aplicada correspondieron al manejo convencional de la zona. En los cultivos de otoño-invierno de los años 2009, 2010 y 2011, se aplicaron concentraciones crecientes de N en el fertiriego (Cuadro 1). El volumen de riego aplicado se basó en la ETc estimada según el modelo de simulación PrHo (Fernández et al., 2001), y la frecuencia fue ajustada según el potencial hídrico del suelo. Las unidades experimentales fueron parcelas de 36 m², distribuidas en un diseño de cuatro bloques completos al azar. Cada parcela tuvo un sistema de riego independiente para la aplicación de los tratamientos de N mediante fertiriego. Las características del invernadero, el sistema de cultivo, el sistema de riego y el manejo agronómico del cultivo fueron descritas detalladamente en Soto et al. (2014).

En todos los cultivos se midieron como variables de respuesta el contenido de humedad volumétrica del suelo, N mineral del suelo, materia seca (MS), rendimiento, extracción de nitrógeno, volumen y concentración de N en el riego aplicado y en drenaje. Además, se estimó la evapotranspiración del cultivo, la cantidad de N aplicado (kg/ha) y lixiviado (kg/ha). Los métodos para la medición y estimación de dichas variables de respuesta se describieron en Soto et al. (2014). En el presente trabajo, para los estudios de casos que ilustran el análisis de regresión propuesto, se seleccionaron las variables de rendimiento comercial de fruto, contenido de humedad volumétrica del suelo, evapotranspiración del cultivo y materia seca, considerando que fueron casos representativos que ilustran la utilidad práctica del análisis de regresión, en escenarios de simulación con resultados diferentes.

Simulación del rendimiento comercial de fruto de tomate

Para la simulación del rendimiento comercial de fruto de tomate, los tratamientos fueron agrupados en dos categorías: de (i) N no limitante, cuando el consumo de N fue menos o igual al nitrógeno disponible en el suelo (NDS) y, (ii) N limitante, cuando el consumo de N fue menor o igual al NDS. Para efectos de ilustrar la aplicación del análisis de regresión en la simulación del rendimiento de fruto, se utilizaron los tratamientos limitantes en N: OI-10-N₁, P-11-N₁ y P-11-N₂ (Cuadro 1).

Simulación del contenido de agua en el suelo (SWC)

Para evaluar la eficiencia del modelo para predecir el SWC (cm³ cm⁻³) en el perfil de 30-60 cm de suelo, se utilizaron los datos de todos los tratamientos (Cuadro 1).

Simulación de la evapotranspiración del cultivo (ET_c) de tomate

Para evaluar la eficiencia del modelo para predecir la ETc, se utilizaron los datos de todos los tratamientos del estudio de Soto et al. (2014) (Cuadro 1).

Simulación de la materia seca (MS) en tomate

Para ilustrar otro escenario con diferentes resultados en la aplicación del análisis de regresión, se utilizó la variable de materia seca (MS) de los tratamientos limitantes en N, OI-10-N₁, P-11-N₁ y P-11-N₂ (Cuadro 1).

Configuración, entradas y validación del modelo

En los cultivos experimentales utilizados para la evaluación del modelo en cultivo de tomate bajo invernadero, se realizó primero una calibración y posteriormente la validación del modelo EU-Rotate_N, usando la versión

1.8. Para cada simulación por tratamiento, el modelo requirió una serie de archivos de entrada (inputs data) que incluyeron información relevante sobre el sitio experimental, las propiedades del suelo, datos del cultivo, manejo del cultivo (riego y fertilización N), contenidos iniciales de humedad y N mineral del suelo, y un archivo con datos de clima de todos los años.

Los parámetros de calibración específicos del cultivo, fueron incluidos en la base de datos del modelo (denominada crop table). Para cada tratamiento se realizó una simulación independiente desde el trasplante hasta la última. En el estudio realizado por Soto et al. (2014), se describieron en detalle la información de las entradas “input data” y parámetros de cultivo “crop table” utilizados en el proceso de validación y calibración del modelo.

Para el análisis estadístico del método de regresión propuesto, se utilizó el paquete estadístico JMP.

Resultados

Simulación del rendimiento comercial de fruto

En la simulación de rendimiento comercial de fruto, las pruebas de hipótesis para intercepto 0 y pendiente 1,0 resultaron no significativas, el RMSE (1,98) fue bajo y por consiguiente el R^2 (0,96) fue muy alto (Figura 1, Cuadros 2 y 3). Esto llevará a límites de confianza para una estimación muy cercanos entre sí, lo que implica que el método de simulación estima lo observado con buena precisión.

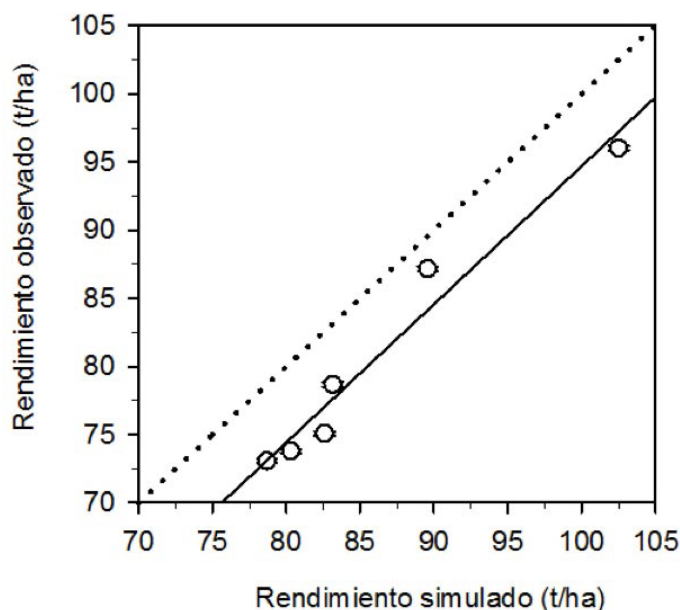


Figura 1. Rendimiento comercial de fruto en cultivo de tomate (*Solanum lycopersicum* L.), bajo invernadero. La línea punteada representa la función identidad. Almería, España. 2011-2012.

Figure 1. Fruit commercial yield in greenhouse tomato (*Solanum lycopersicum* L.) crop. The dotted line represents the identity function. Almeria, Spain 2011-2012.

Cuadro 2. Resumen del ajuste entre valores simulados y observados del rendimiento comercial de fruto (t/ha) en cultivo de tomate (*Solanum lycopersicum* L.), bajo invernadero. Almería, España. 2011-2012.

Table 2. Summary of the adjustment between simulated and observed values of fruit commercial yield (t/ha) in greenhouse tomato (*Solanum lycopersicum* L.) crop. Almeria, Spain. 2011-2012.

Término	Estimación
R Cuadrado	0,956
R Cuadrado ajustado	0,947
Raíz cuadrada del cuadrado medio de error	1,980
Respuesta promedio	79,846
Observaciones	7

Cuadro 3. Parámetros estimados para el ajuste entre valores simulados y observados de rendimiento comercial de fruto (t/ha) en cultivo de tomate (*Solanum lycopersicum* L.), bajo invernadero. Almería, España. 2011-2012.

Table 3. Estimated parameters for the adjustment between simulated and observed values of fruit commercial yield (t/ha) in greenhouse tomato (*Solanum lycopersicum* L.), crops. Almeria, Spain. 2011-2012.

Término	Estimación	Error estándar	t	Prob> t
Intercepto (b_0)	-8,165746	8,490776	-0,96	0,3804
Rendimiento N L Simulado (b_1)	1,0279251	0,098782	10,41	0,0001*

En este caso, destaca el hecho de que las hipótesis que prueba el paquete de análisis para β_0 y β_1 fueron:

$$H_0: \beta_0 = 0 \quad H_0: \beta_1 = 0$$

En el caso β_1 (la pendiente), lo que se quiso fue probar la hipótesis de que $H_0: \beta_1 = 1,0$, contra la alternativa de que es diferente de 1,0. Esto puede hacerse con facilidad definiendo el valor del tRatio (t de Student) de la siguiente manera (ecuación 13):

$$t = \frac{(b_1 - 1)}{\text{stdError}(b_1)} = \frac{(1,0279251 - 1)}{0,098782} = 0,2826 \quad (13)$$

La “t” se calcularon restando 2 al número total de observaciones, donde en este caso fueron 5.

La $\text{Prob}>|0,2826|$ con 5 grados de libertad fue 0,7888, lo que implica que no se rechace la hipótesis nula planteada. Así, el apartado de “parámetros estimados” en la salida de JMP aparecería de la siguiente manera (Cuadro 4):

Cuadro 4. Parámetros estimados para el ajuste entre valores simulados y observados de rendimiento comercial de fruto en cultivo de tomate (*Solanum lycopersicum* L.), bajo invernadero. Almería, España. 2011-2012.

Table 4. Estimated parameters for the adjustment between simulated and observed values of fruit commercial yield in greenhouse tomato (*Solanum lycopersicum* L.), crop. Almeria, Spain. 2011-2012.

Término	Estimación	Error estándar	t	Prob> t
Intercepto (b_0)	-8,165746	8,491	-0,96	0,3804
Rendimiento simulado (b_1)	1,0279251	0,099	0,28	0,7888

En este caso, el intercepto no fue significativamente diferente de 0 ($p=0,3804$) y la pendiente no fue significativamente diferente de 1,0 ($p=0,7888$).

Los resultados del ajuste para cada uno de los valores simulados se muestran en el Cuadro 5. Los límites de confianza pueden ser utilizados para valorar la utilidad práctica del modelo de simulación. Por ejemplo, a un valor simulado igual a $83,14 \text{ t}\cdot\text{ha}^{-1}$ corresponde un valor predicho por la ecuación de $77,3 \text{ t}\cdot\text{ha}^{-1}$. Los límites de confianza indicaron que hubo una confianza del 95 % de que el valor verdadero del promedio por hectárea del rendimiento comercial de fruto se encuentre entre $75,27$ y $79,32 \text{ t}\cdot\text{ha}^{-1}$. En el peor de los casos se esperaría un rendimiento de $75,27 \text{ t}\cdot\text{ha}^{-1}$. El valor absoluto del error de estimación fue, en este caso, de 2,02 que en términos relativos representó un 2,6 % del valor predicho (Cuadro 5).

Cuadro 5. Resultados del ajuste entre valores observados y simulados para el rendimiento comercial de fruto (t/ha) en cultivo de tomate (*Solanum lycopersicum* L.), bajo invernadero. Almería, España. 2011-2012.

Table 5. Results of the adjustment between simulated and observed values for the commercial performance of the fruit (t/ha) in greenhouse tomato (*Solanum lycopersicum* L.), crop. Almeria, Spain. 2011-2012.

Observado	Simulado	Predicho por la ecuación	Residuo	Límite inferior	Límite superior
75,09	82,57	76,71	-1,62	74,64	78,78
73,80	80,31	74,39	-0,59	72,04	76,74
96,02	102,50	97,20	-1,18	92,50	101,89
87,16	89,57	83,91	3,25	81,74	86,08
78,68	83,14	77,30	1,38	75,27	79,32
73,08	78,68	72,71	0,37	70,10	75,32
75,09	82,57	76,71	-1,62	74,64	78,78

Simulación del contenido de agua en el suelo (SWC)

Las pruebas de hipótesis para intercepto 0 y pendiente 1,0 resultaron no significativas ($P>0,05$), pero el RMSE fue alto y por consiguiente se obtuvo un R^2 muy bajo. Esto llevará a límites de confianza para una estimación promedio (ecuación 1) muy anchos, lo que implica que el método de simulación estimó lo observado con muy poca precisión (Figura 2, Cuadros 6 y 7).

Para un simulado de 0,19, que se encuentra cerca del punto medio de la distribución al que corresponde un valor predicho de 0,184 (Cuadro 8), el error de estimación fue $\pm 0,075$, que en términos relativos fue un 4,1 % del estimado. Cuando el valor se aleja del centro de la distribución, el error aumenta. Por ejemplo, para un simulado de 0,240 el error es de $\pm 0,11$ que representa un 4,9 % del valor predicho. Como se indicó anteriormente, aunque se cumple la condición de pendiente 1,0 e intercepto 0, un R^2 bajo y un RMSE alto implican una menor precisión en la estimación.

Simulación de la evapotranspiración del cultivo (ET_c)

En la simulación de la evapotranspiración del cultivo (ET_c), el intercepto fue significativamente diferente de 0 y la pendiente fue significativamente diferente de 1,0, pero el tamaño de la diferencia en ambos casos fue irrelevante (Figura 3). La significancia indica que lo obtenido para la muestra fue generalizable a la población, sin

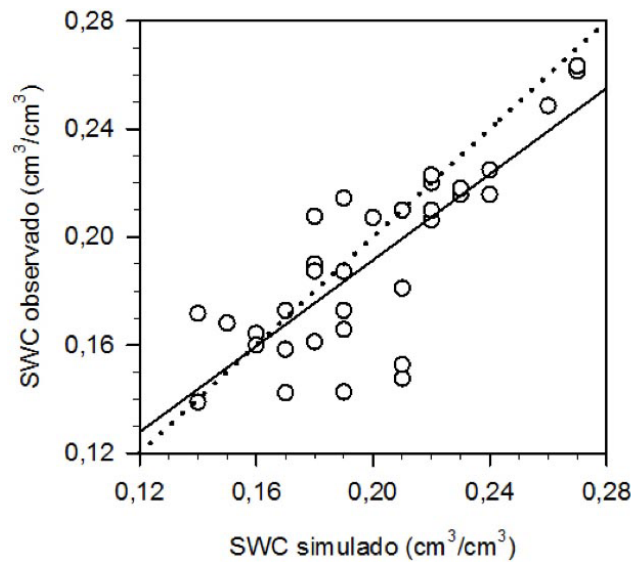


Figura 2. Comparación entre valores medidos y simulados del contenido de humedad volumétrica del suelo (SWC) en el perfil de 30-60 cm, en cultivos de tomate (*Solanum lycopersicum* L.) bajo invernadero. La línea punteada es la función identidad. Almería, España. 2011-2012.

Figure 2. Comparison between measured and simulated values of soil water content (SWC) on the 30-60 cm profile in greenhouse tomato (*Solanum lycopersicum* L.) crop. The dotted line is the identity function. Almeria, Spain. 2011-2012.

Cuadro 6. Resumen de ajuste para valores medidos y simulados de humedad volumétrica en el perfil de 30-60 cm del suelo, en cultivo de tomate (*Solanum lycopersicum* L.) bajo invernadero. Almería, España. 2011-2012.

Table 6. Adjustment summary of measured and simulated values of volumetric soil moisture at 30-60 cm profile, in greenhouse tomato (*Solanum lycopersicum* L.) crop. Almeria, Spain. 2011-2012.

Término	Estimación
R cuadrado	0,652
R cuadrado ajustado	0,641
Raíz cuadrada del cuadrado medio de error	0,020
Respuesta promedio	0,192
Observaciones	34

Cuadro 7. Parámetros estimados para el ajuste entre valores medidos y simulados de humedad volumétrica (cm^3/cm^3) en el perfil de 30-60 cm del suelo, en cultivo de tomate (*Solanum lycopersicum* L.), bajo invernadero. Almería, España. 2011-2012.

Table 7. Estimated parameters for the adjustment between measured and simulated values of volumetric soil moisture (cm^3/cm^3) at 30-60 cm profile, in greenhouse tomato (*Solanum lycopersicum* L.) crop. Almeria, Spain. 2011-2012.

Término	Estimación	Error estándar	t	Prob> t
Intercepto (b_0)	0,031	0,021	1,49	0,1455
Humedad Vol. 30-60 cm simulado (b_1)	0,802	0,104	1,91 ¹	<0,066 ¹

¹Para la prueba de hipótesis de $\beta_1=1$ / For the hypothesis test of $\beta_1=1$.

Cuadro 8. Resultados del ajuste entre valores medidos y simulados de humedad volumétrica (cm^3/cm^3) en el perfil de 30-60 cm del suelo, en cultivo de tomate (*Solanum lycopersicum* L.) bajo invernadero. Almería, España. 2011-2012.

Table 8. Results of the adjustment between measured and simulated values of volumetric soil moisture (cm^3/cm^3) on the 30-60 cm profile, in greenhouse tomato (*Solanum lycopersicum* L.) crop. Almeria, Spain. 2011-2012.

Observado	Simulado	Predicho	Residuo	Límite inferior	Límite superior
0,216	0,240	0,224	-0,008	0,213	0,235
0,148	0,210	0,200	-0,052	0,192	0,207
0,158	0,170	0,168	-0,009	0,158	0,177
0,216	0,230	0,216	0,000	0,206	0,225
0,207	0,200	0,192	0,015	0,185	0,199
0,172	0,140	0,144	0,028	0,129	0,158
0,225	0,240	0,224	0,001	0,213	0,235
0,143	0,190	0,184	-0,041	0,176	0,191
0,166	0,190	0,184	-0,018	0,176	0,191
0,206	0,220	0,208	-0,002	0,200	0,216
0,161	0,180	0,176	-0,015	0,167	0,184
0,214	0,190	0,184	0,030	0,176	0,191
0,261	0,270	0,248	0,013	0,232	0,264
0,210	0,220	0,208	0,002	0,200	0,216
0,173	0,170	0,168	0,005	0,158	0,177
0,187	0,190	0,184	0,003	0,176	0,191
0,263	0,270	0,248	0,015	0,232	0,264
0,210	0,210	0,200	0,010	0,192	0,207
0,173	0,190	0,184	-0,011	0,176	0,191
0,187	0,180	0,176	0,012	0,167	0,184
0,223	0,220	0,208	0,015	0,200	0,216
0,142	0,170	0,168	-0,025	0,158	0,177
0,208	0,180	0,176	0,032	0,167	0,184
0,218	0,230	0,216	0,002	0,206	0,225
0,153	0,210	0,200	-0,047	0,192	0,207
0,168	0,150	0,152	0,016	0,139	0,164
0,139	0,140	0,144	-0,005	0,129	0,158
0,249	0,260	0,240	0,009	0,226	0,254
0,181	0,210	0,200	-0,019	0,192	0,207
0,164	0,160	0,160	0,005	0,149	0,171
0,160	0,160	0,160	0,000	0,149	0,171
0,220	0,220	0,208	0,012	0,200	0,216

embargo, fue provocada por un RMSE extremadamente bajo, lo cual es muy deseable, y no significa que realmente el intercepto sea diferente de cero y la pendiente sea diferente de uno. Este resultado se vio, además, apoyado por un R^2 sustancialmente alto (0,982) (Cuadro 9).

El valor muestral del intercepto fue de -3,9, la cual es una diferencia irrelevante en el orden de magnitud de la variable que varía entre 0 y 350 (Cuadro 10). El valor muestral de la pendiente también mostró una diferencia

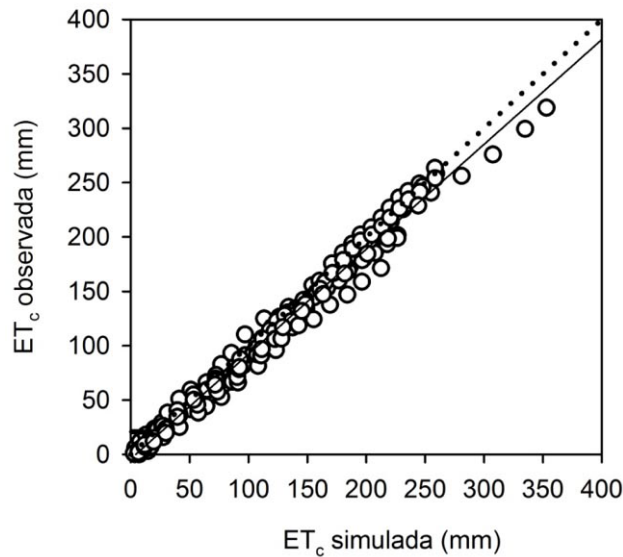


Figura 3. Comparación entre valores medidos y simulados de la evapotranspiración (ET_c) en cultivo de tomate (*Solanum lycopersicum* L.), bajo invernadero. La línea punteada es la función identidad. Almería, España. 2011-2012.

Figure 3. Comparison between measured and simulated values of evapotranspiration (ET_c) in greenhouse tomato (*Solanum lycopersicum* L.) crop. The dotted line is the identity function. Almeria, Spain. 2011-2012.

Cuadro 9. Resumen de ajuste para valores medidos y simulados de evapotranspiración de cultivos (mm) de tomate (*Solanum lycopersicum* L.), en suelo bajo invernadero. Almería, España. 2011-2012.

Table 9. Adjustment summary between measured and simulated values of evapotranspiration (mm) of soil grown greenhouse tomato (*Solanum lycopersicum* L.). Almeria, Spain. 2011-2012.

Término	Estimación
R cuadrado	0,982
R cuadrado ajustado	0,981
Raíz cuadrada del cuadrado medio de error	10,490
Respuesta promedio	114,809
Observaciones	209

irrelevante con el valor de 1,0 que se desea para ella. Nótese el valor de 0,009 para el error estándar (Cuadro 10), ese valor tan bajo, muy deseado en todos los procesos de medición, fue el que hizo que ambas hipótesis se rechacen en estas condiciones.

Cuadro 10. Parámetros estimados para el ajuste entre valores medidos y simulados de evapotranspiración (mm) en cultivo de tomate (*Solanum lycopersicum* L.) bajo invernadero. Almería, España. 2011-2012.

Table 10. Estimate parameters for the adjustment between measured and simulated values of evapotranspiration (mm) in greenhouse tomato (*Solanum lycopersicum* L.) crop. Almeria, Spain. 2011-2012.

Término	Estimación	Error estándar	t	Prob> t
Intercepto (b_0)	-3,896	1,331	-2,93	0,0038*
Evapotranspiración simulado (b_1)	0,963	0,009	4,05 ¹	<0,0001 ¹

¹Para la prueba de hipótesis de $\beta_1=1$ / For the hypothesis test of $\beta_1=1$.

Simulación de la materia seca (MS) en tomate

En la simulación de la materia seca (MS), la línea de ajuste no fue 1:1 porque la pendiente fue mayor que 1,0. También en este caso se dio un RMSE bajo, lo que implica un R^2 alto, pero las diferencias entre lo simulado por el modelo y lo observado crecen conforme los valores de la variable aumentan, lo que hace que la simulación no tenga validez alguna. Los resultados se muestran en la Figura 4 y en los Cuadros 11 y 12.

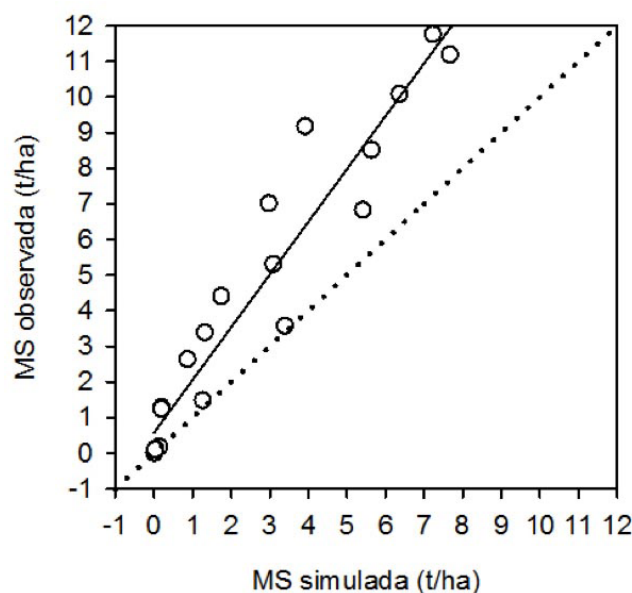


Figura 4. Comparación entre valores medidos y simulados de materia seca (MS) en cultivo de tomate (*Solanum lycopersicum* L.), bajo invernadero. La línea punteada fue la función identidad. Almería, España. 2011-2012.

Figure 4. Comparison between measured and simulated values of dry matter (MS) in greenhouse tomato (*Solanum lycopersicum* L.) crop. The dotted line was the identity function. Almeria, Spain. 2011-2012.

Cuadro 11. Resumen de ajuste para valores medidos y simulados de materia seca (t/ha) en cultivos de tomate (*Solanum lycopersicum* L.), en suelo bajo invernadero. Almería, España. 2011-2012.

Table 11. Adjustment summary of measured and simulated values of dry matter (t/ha) in greenhouse tomato (*Solanum lycopersicum* L.) crop. Almeria, Spain. 2011-2012.

Término	Estimación
R cuadrado	0,920
R cuadrado ajustado	0,916
Raíz cuadrada del cuadrado medio de error	1,175
Respuesta promedio	4,208
Observaciones	21

Cuadro 12. Parámetros estimados para el ajuste entre valores medidos y simulados de materia seca (t/ha) en cultivo de tomate (*Solanum lycopersicum* L.) bajo invernadero. Almería, España. 2011-2012.

Table 12. Estimated parameters for the adjustment between measured and simulated values of dry matter (t/ha) in greenhouse tomato (*Solanum lycopersicum* L.) crop. Almeria, Spain. 2011-2012.

Término	Estimación	Error estándar	t	Prob> t
Intercepto	0,578	0,354	1,63	0,1196
Materia seca simulada	1,481	0,010	4,82	<0,0001 ¹

¹Para la prueba de hipótesis de $\beta_1=1$ / For the hypothesis test of $\beta_1=1$.

Discusión

Todas las medidas que han sido propuestas para evaluar el modelo de simulación basado en las diferencias entre valores observados y simulados, calculan dichas diferencias para una muestra, con lo cual proveen información del desempeño del modelo de simulación para esa muestra específica, pero si su resultado se generaliza a la población, no se dispone de una medida del error asociado a esa generalización. Por tanto, su utilidad es limitada.

Habitualmente, cuando se evalúa el desempeño de modelos de simulación, la mayoría de autores utilizan métodos “por costumbre”, evitan los detalles metodológicos básicos, y utilizan terminología y simbología que conlleva a ambigüedades y confusión. No se cuestiona si al utilizar un estadístico dado, este podría ser más útil cuando se compara la capacidad de simulación entre dos modelos diferentes, o cuando se compara lo observado con lo simulado en un mismo modelo.

Otros estadísticos, como por ejemplo el error relativo, pueden verse afectados por la magnitud de los valores, por valores extremos y el número de observaciones. En otros casos, se utiliza el RMSE indiscriminadamente, sin considerar si se aplica en el contexto del análisis de diferencias o en el contexto del análisis de regresión.

Los casos presentados ilustraron el uso correcto del análisis de regresión para la validación de modelos de simulación en agricultura. En todos ellos la variable predictora (X) fue el valor determinado por la simulación y la variable respuesta el valor observado (Y), contrario a lo que utilizaron Yang et al. (2014) y Willmott (1982). Cada uno de ellos fue de diferente índole y se mostraron resultados para los ajustes de regresión que permitieron, en conjunto, decidir si el modelo fue válido o no. Esto justifica el por qué en este trabajo, reforzado con los ejemplos ilustrados, el análisis de regresión propuesto se reconoce como la mejor opción para evaluar la capacidad de

predicción de un modelo específico o en particular. En el análisis de regresión deben considerarse conjuntamente el R^2 (porcentaje de la variancia de Y explicado por el modelo), el RMSE (el cuadrado medio de error), los valores y significancia del coeficiente de regresión y la constante del modelo, y los límites de confianza para las estimaciones.

En la propuesta de Reckhow et al. (1990), Mayer et al. (1994) y Flavelle (1992), sobre la regresión como técnica de validación, señalaron su limitación cuando se incumplen los supuestos básicos. Es importante la evaluación del cumplimiento de los supuestos básicos de normalidad de los errores y homoscedasticidad, y el uso de cuadrados mínimos ponderados en el caso de que esta última no se satisfaga.

Finalmente, la utilización de un método estadístico apropiado, como es el análisis de regresión utilizado en forma correcta, junto con la experiencia profesional en un área específica, parece ser la combinación más adecuada para evaluar la capacidad de simulación de un modelo y llegar así a inferencias o conclusiones válidas.

Conclusiones

Se concluye que el análisis de regresión, utilizado correctamente, es el método más adecuado para evaluar el desempeño de un modelo de simulación en agricultura, sin embargo, debe considerarse conjuntamente con el R^2 , el RMSE, los valores y la significancia del coeficiente de regresión, la constante del modelo y los límites de confianza para las estimaciones.

Estadísticos tales como el ER, el d y la E, son habitualmente usados de forma errónea para evaluar el desempeño de un modelo de simulación agrícola como tal, ya que son más recomendables para la comparación del desempeño entre diferentes modelos de simulación. El RMSE es habitualmente utilizado para evaluar diferencias entre valores observados y simulados, sin discriminar que este es diferente al RMSE de una regresión.

Literatura citada

- Cannavo, P.R., S. Parnaudeau, and V. Reau. 2008. Modeling N dynamics to assess environmental impacts of cropped soils. *Adv. Agron.* 97:131-174. doi:10.1016/S0065-2113(07)00004-1
- Doltra, J., and P. Muñoz. 2010. Simulation of nitrogen leaching from a fertigated crop rotation in a Mediterranean climate using the EU-Rotate_N and Hydrus-2D models. *Agric. Water Manag.* 97:277-285. doi:10.1016/j.agwat.2009.09.019
- Draper, N., and H. Smith. 1998. *Applied regression analysis*. 3rd ed. Wiley Interscience, NY, USA.
- Fernández, M., F. Orgaz, E. Fereres, J.C. López, A. Céspedes, J. Pérez, S. Bonachela, y M. Gallardo. 2001. Programación del riego de cultivos hortícolas bajo invernadero en el sudeste español. Caja Rural, Almería, ESP.
- Flavelle, P. 1992. A quantitative measure of model validation and its potential use for regulatory purpose. *Adv. Water Resour.* 15:5-13. doi:10.1016/0309-1708(92)90028-Z
- Fox, D.G. 1981. Judging air quality model performance: A summary of the AMS workshop on dispersion model performance. *Bull. Am. Meteorol. Soc.* 62:599-609.
- Galloway, J., F. Dentener, D. Capone, E. Boyer, R. Howarth, S. Seitzinger, G. Asner, C. Cleveland, P. Green, E. Holland, D. Karl, F. Michaels, J. Porter, A. Townsend, and C. Voronarty. 2004. The global nitrogen cycle: Past, present and future. *Biogeochemistry* 70:153-226. doi:10.1007/s10533-004-0370-0
- Guo, R., C. Nendel, C. Rahn, C. Jiang, and Q. Chen. 2010. Tracking nitrogen losses in a greenhouse crop rotation experiment in North China using the EU-Rotate_N simulation model. *Environ. Pollut.* 158:2218-2229. doi:10.1016/j.envpol.2010.02.014

- Mayer, D.G., M.A. Stuart, and A.J. Swain. 1994. Regression of real word data on model output: an appropriate overall test of validity. *Agric. Syst.* 45:93-104. doi:10.1016/S0308-521X(94)90282-8
- Meinke, H., P. Struik, J. Vos, and W. Van-der-Werf. 2008. Modelling that bridges scales and connects disciplines. In: H. Van-Keulen et al., editors, 40 years theory and model at Wageningen UR. Wageningen University, and Research Centre, Wageningen, NED. p. 37.
- Mendenhall, W., R. Scheaffer, y D. Wackerly. 1986. Estadística matemática con aplicaciones. 2a ed. Grupo Editorial Iberoamérica, Andalucía, ESP.
- Nash, J.E., and J.V Sutcliffe. 1970. River Flow forecasting through conceptual models. Part I-A discussion of principles. *J. Hydrol.* 10:282-290. doi:10.1016/0022-1694(70)90255-6
- Rahn, C.R., K. Zhang, R. Lillywhite, C. Ramos, J. Doltra, J.M. de-Paz, H. Riley, M. Fink, C. Nendel, K. Thorup-Kristensen, A. Pedersen, F. Piro, A. Venezia, C. Firth, U. Schmutz, F. Rayns, and K. Strohmeyer. 2010. Eu-Rotate_N - a decision support system - to predict environmental and economic consequences of the management of nitrogen fertiliser in crop rotations. *Europ. J. Hort. Sci.* 75(1):20-32.
- Reckhow, K.H., J.T. Clements, and R.C. Dodds. 1990. Statistical evaluation of mechanistic water-quality models. *J. Environ. Eng.* 116:250-268. doi:10.1061/(ASCE)0733-9372(1990)116:2(250)
- Soto, F., M. Gallardo, C. Giménez, T. Peña-Fleitas, and R.B. Thompson. 2014. Simulation of tomato growth, water and N dynamics using the EU-Rotate_N model in Mediterranean greenhouses with drip irrigation and fertigation. *Agric. Water Manag.* 132:46-59. doi:10.1016/j.agwat.2013.10.002
- Stockle, C.O., J. Kjelgaard, and G. Bellocchi. 2004. Evaluation of estimated weather data for calculating Penman-Monteith reference crop evapotranspiration. *Irrig. Sci.* 23:39-46. doi:10.1007/s00271-004-0091-0
- Sun, Y., K. Hu, K. Zhang, L. Jiang, and Y. Xu. 2012. Simulation of nitrogen fate for greenhouse cucumber grown under different water and fertilizer management using the EU-Rotate N model. *Agric. Water Manag.* 112:21-32. doi:10.1016/j.agwat.2012.06.001
- Thornley, J.H., and I.M. Johnson. 2000. Plant and crop modelling: A mathematical approach to plant and crop physiology. The Blackburn Press, NJ, USA.
- Tilman, D., K.G. Cassman, P.A. Matson, R. Naylor, and S. Polasky. 2002. Agricultural sustainability and intensive production practices. *Nature* 418:671-677. doi:10.1038/nature01014
- Willmott, C.J. 1982. Some comments on the evaluation of model performance. *Bull. Am. Meteorol. Soc.* 63:1309-1313. doi:10.1175/1520-0477(1982)063<1309:SCOTEO>2.0.CO;2
- Yang, J., D.J. Greenwood, D.L. Rowell, G.A. Wadsworth, and I.G. Burns. 2000. Statistical methods for evaluating a crop nitrogen simulation model, N_ABLE. *Agric. Syst.* 64:37-53. doi:10.1016/S0308-521X(00)00010-X
- Yang, J.M., J.Y. Yang, S. Liu, and G. Hoogenboom. 2014. An evaluation of the statistical methods for testing the performance of crop models with observed data. *Agric. Syst.* 127:81-89. doi:10.1016/j.agsy.2014.01.008